DEVIATION INEQUALITIES FOR BIFURCATING MARKOV CHAINS ON GALTON-WATSON TREE

S.VALÈRE BITSEKI PENDA

Abstract. We provide deviation inequalities for properly normalized sums of bifurcating Markov chains on Galton-Watson tree. These processes are extension of bifurcating Markov chains (which was introduced by Guyon to detect cellular aging from cell lineage) in case the index set is a binary Galton-Watson process. As application, we derive deviation inequalities for the least-squares estimator of autoregressive parameters of bifurcating autoregressive processes with missing data. These processes allow, in case of cell division, to take into account the cell's death. The results are obtained under an uniform geometric ergodicity assumption of an embedded Markov chain.

Key words: Bifurcating Markov chains, Galton-Watson processes, ergodicity, deviation inequalities, first order bifurcating autoregressive process with missing data, cellular aging.

AMS 2000 subject classifications. Primary 60E15, 60J80; secondary 60J10.

1. Introduction

Bifurcating Markov chains (BMC) on Galton-Watson (GW) tree are an extension of BMC to GW tree data. They were introduced by Delmas and Marsalle [12] in order to take into account the death of individuals in the Escherichia coli's (E.coli) reproduction model. E.coli is a rod-shaped bacterium which reproduces by dividing in the middle, thus producing two cells. One which has the new pole of the mother and that we call new pole progeny cell, and the other which has the old pole of the mother and that we call old pole progeny cell. In fact, each daughter cell has two poles. One which is new (new pole) and the other which already existed (old pole). The age of a cell is given by the age of its old pole (i.e the number of generations in the past of the cell before the old pole was produced).

Guyon & Al [15] proposed the following linear Gaussian model to describe the evolution of the growth rate of the population of cells derived from an initial individual:

$$\mathcal{L}(X_1) = \nu, \qquad \text{and} \qquad \forall n \geq 1, \quad \begin{cases} X_{2n} = \alpha_0 X_n + \beta_0 + \varepsilon_{2n} \\ \\ X_{2n+1} = \alpha_1 X_n + \beta_1 + \varepsilon_{2n+1}, \end{cases} \tag{1.1}$$

where X_n is the growth rate of individual n, n is the mother of $2n$ (the new pole progeny cell) and $2n+1$ (the old pole progeny cell), ν is a distribution probability on $\mathbb{R}$, $\alpha_0, \alpha_1 \in (-1,1)$; $\beta_0, \beta_1 \in \mathbb{R}$ and $\big((\varepsilon_{2n}, \varepsilon_{2n+1}), n \geq 1\big)$ forms a sequence of i.i.d bivariate random variables with law $\mathcal{N}_2(0,\Gamma)$, where

$$\Gamma = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad \sigma^2 > 0, \quad \rho \in (-1,1). \tag{1.2}$$

The processes (X_n) defined by (1.1) are typical example of BMC which are called the first order bifurcating autoregressive processes (BAR(1)). The BAR(1) processes are an adaptation

of autoregressive processes, when the data have a binary tree structure (see Figure 1). They were first introduced by Cowan and Staudte [9] for cell lineage data where each individual in one generation gives rise to two offspring in the next generation.

In [14], Guyon, using the theory of BMC, gave laws of large numbers and central limit theorem for the least-squares estimator $\widehat{\theta}^r = (\widehat{\alpha}_0^r, \widehat{\beta}_0^r, \widehat{\alpha}_1^r, \widehat{\beta}_1^r)$ of the 4-dimensional parameter $\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1)$. He has also built some statistical tests which allow to test if the model is symmetric or not, and if the new pole and the old pole populations are even distinct in mean. This allowed him to conclude a statistical evidence in aging in E. Coli. Let us also mention [6], where Bercu & Al. using the martingale approach give asymptotic analysis of the least squares estimator of the unknown parameters of a general asymmetric pth-order BAR processes.

However, in the BMC model presented by Guyon, cells are assumed to never die (a death corresponds to no more division). To take into account cells's death, Delmas and Marsalle [12], instead of a regular binary tree, used a binary GW tree to label cells. In the sequel, we will introduce the model which allowed them to study the behavior of the growth rate of cells, taking into account their possible death.

1.1. **The model.** Let $\mathbb{T}$ be a binary regular tree in which each vertex is seen as a positive integer different from 0, see Figure 1. For $r \in \mathbb{N}$, let

$$\mathbb{G}_r = \left\{2^r, 2^r+1, \cdots, 2^{r+1}-1\right\}, \quad \mathbb{T}_r = \bigcup_{q=0}^{r} \mathbb{G}_q,$$

which denote respectively the r-th column and the first $(r+1)$ columns of the tree. Then, the cardinality $|\mathbb{G}_r|$ of $\mathbb{G}_r$ is 2^r and that of $\mathbb{T}_r$ is $|\mathbb{T}_r| = 2^{r+1}-1$. A column of a given integer n is $\mathbb{G}_{r_n}$ with $r_n = \lfloor \log_2 n \rfloor$, where $\lfloor x \rfloor$ denotes the integer part of the real number x.

The genealogy of the cells is described by this tree. In the sequel we will thus see $\mathbb{T}$ as a given population. Then the vertex n, the column $\mathbb{G}_r$ and the first $(r+1)$ columns $\mathbb{T}_r$ designate respectively individual n, the r-th generation and the first $(r+1)$ generations. The initial individual is denoted 1. The model proposed by Delmas and Marsalle [12] is defined as follows. The growth rate of cell n is X_n.

- With probability $p_{1,0}$, n gives birth to two cells $2n$ and $2n+1$ with both divide. The growth rate of the daughters X_{2n} and X_{2n+1} are then linked to the mother's one through auto-regressive equations (1.1).
- With probability p_0, only the new pole $2n$ divides. Its growth rate X_{2n} is linked to its mother's one X_n through the relation
$$X_{2n} = \alpha_0' X_n + \beta_0' + \varepsilon_{2n}', \tag{1.3}$$
where $\alpha_0' \in (-1,1)$, $\beta_0' \in \mathbb{R}$ and $(\varepsilon_{2n}', n \in \mathbb{T})$ is a sequence of independent centered Gaussian random variables with variance $\sigma_0^2 > 0$.
- With probability p_1, only the old pole $2n+1$ divides. Its growth rate X_{2n+1} is linked to its mother's one X_n through the relation
$$X_{2n+1} = \alpha_1' X_n + \beta_1' + \varepsilon_{2n+1}', \tag{1.4}$$
where $\alpha_1' \in (-1,1)$, $\beta_1' \in \mathbb{R}$ and $(\varepsilon_{2n+1}', n \in \mathbb{T})$ is a sequence of independent centered Gaussian random variables with variance $\sigma_1^2 > 0$.

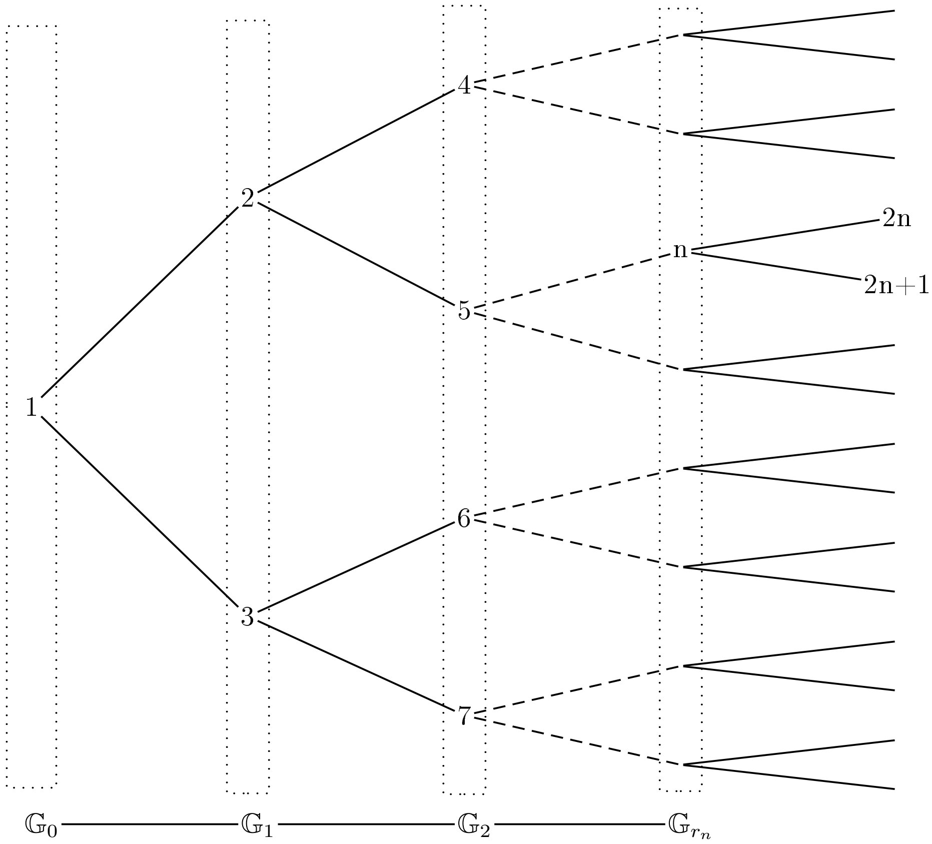

FIGURE 1. The binary tree $\mathbb{T}$

- With probability $1 - p_{1,0} - p_1 - p_0$, which is non-negative, n gives birth to two cells which do not divide.
- The sequences $((\varepsilon_{2n}, \varepsilon_{2n+1}), n \in \mathbb{T})$, $(\varepsilon'_{2n}, n \in \mathbb{T})$ and $(\varepsilon'_{2n+1}, n \in \mathbb{T})$ are independent.

The process (X_n) described above is a typical example of BMC on GW tree. In [11], this process is called bifurcating autoregressive process (BAR) with missing data. It is an extension of bifurcating autoregressive process when the data have a binary GW tree structure, see figure 2 for example of binary GW tree. Indeed, one can assume that the cells which do not divide and those which do not exist are missing or dead.

In [12], Delmas and Marsalle using their results for BMC on GW tree, gave laws of large numbers and central limit theorem for the maximum likelihood estimator of the parameter

$$\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1, \alpha'_0, \beta'_0, \alpha'_1, \beta'_1). \tag{1.5}$$

In this paper, we will give deviation inequalities for the least squares estimator of the parameter θ, in case the noise sequence and the initial state X_1 take their values in a compact set. Note that this implies that the BAR process with missing data describes above also take their values in compact set. These deviation inequalities are important for a rigorous non asymptotic statistical study. Indeed, when the sample size is insufficient to apply limit

theorems, they allow for example to estimate the errors in the estimation of unknown parameters. Furthermore, these inequalities allow to get a rate of convergence in the laws of large numbers, and this permit, for example, to build non-asymptotic confidence intervals.

We are now going to give a rigorous definition of BMC on GW tree. We refer to [12] for more details.

1.2. **Definitions.** For an individual $n \in \mathbb{T}$, we are interested in the quantity X_n (it may be the weight, the growth rate,$\cdots$) with values in the metric space S endowed with its Borel σ-field $\mathcal{S}$.

Definition 1.1 ($\mathbb{T}$-transition probability, see ([14])). *We call $\mathbb{T}$-transition probability any mappings $P : S \times \mathcal{S}^2 \to [0, 1]$ such that*

- *$P(., A)$ is measurable for all $A \in \mathcal{S}^2$,*
- *$P(x, .)$ is a probability measure on $(S^2, \mathcal{S}^2)$ for all $x \in S$.*

For $p \geq 1$, we denote by $\mathcal{B}(S^p)$(resp. $\mathcal{B}_b(S^p)$, $\mathcal{C}(S^p)$, $\mathcal{C}_b(S^p)$) the set of all $\mathcal{S}^p$-measurable (resp. $\mathcal{S}^p$-measurable and bounded, continuous, continuous and bounded) mapping $f : S^p \to \mathbb{R}$. For $f \in \mathcal{B}(S^3)$, when it is defined, we denote by $Pf \in \mathcal{B}(S)$ the function

$$x \mapsto Pf(x) = \int_{S^2} f(x, y, z)P(x, dy, dz).$$

Definition 1.2 (Bifurcating Markov Chains, see ([14])). *Let $(X_n, n \in \mathbb{T})$ be a family of S-valued random variables defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_r, r \in \mathbb{N}), \mathbb{P})$. Let ν be a probability on $(S, \mathcal{S})$ and P be a $\mathbb{T}$-transition probability. We say that $(X_n, n \in \mathbb{T})$ is a $(\mathcal{F}_r)$-bifurcating Markov chain with initial distribution ν and $\mathbb{T}$-transition probability P if*

- *X_n is $\mathcal{F}_{r_n}$-measurable for all $n \in \mathbb{T}$,*
- *$\mathcal{L}(X_1) = \nu$,*
- *for all $r \in \mathbb{N}$ and for all family $(f_n, n \in \mathbb{G}_r) \subseteq \mathcal{B}_b(S^3)$*

$$\mathbb{E}\left[\prod_{n\in\mathbb{G}_r} f_n(X_n, X_{2n}, X_{2n+1})\Big|\mathcal{F}_r\right] = \prod_{n\in\mathbb{G}_r} Pf_n(X_n).$$

Now, we add a cemetery point to S, ∂. Let $\bar{S} = S \cup \{\partial\}$, and $\bar{\mathcal{S}}$ be the $\sigma-$field generated by $\mathcal{S}$ and $\{\partial\}$. In the previous biological framework, S corresponds to the state space of the quantities related to living cells, and ∂ is the default value for dead cells. Let P^* be a $\mathbb{T}$-transition probability defined on $\bar{S} \times \bar{\mathcal{S}}$ such that

$$P^*(\partial, \{(\partial, \partial)\}) = 1. \tag{1.6}$$

In the previous biological framework, (1.6) means that no dead cell can give birth to a living cell. We denote by P_0^* and P_1^* the restriction of the first and the second marginal of P^* to S, that is:

$$P_0^* = P^*\left(\cdot, \left(\cdot \bigcap S\right) \times \bar{S}\right) \quad \text{and} \quad P_1^* = P^*\left(\cdot, \bar{S} \times \left(\cdot \bigcap S\right)\right).$$

Definition 1.3 (BMC on GW tree, see [12]). *Let $X = (X_n, n \in \mathbb{T})$ be a P^*-BMC on $(\bar{S}, \bar{\mathcal{S}})$, with P^* satisfying (1.6). We call $(X_n, n \in \mathbb{T}^*)$, with $\mathbb{T}^* = \{n \in \mathbb{T} : X_n \neq \partial\}$, a BMC on GW tree. The P^*-BMC is said spatially homogeneous if $p_{1,0} = P^*(x, S \times S)$, $p_0 = P^*(x, S \times \{\partial\})$, and $p_1 = P^*(x, \{\partial\} \times S)$ do not depend on $x \in S$. A spatially homogeneous P^*-BMC is said super-critical if $m > 1$, where $m = 2p_{1,0} + p_1 + p_0$.*

We denote by $(Y_n, n \in \mathbb{N})$ the Markov chain on S with $Y_0 = X_1$ and transition probability $Q = \frac{1}{m}(P_0^* + P_1^*)$.

Remark 1.4.
- The name BMC on GW tree comes from the fact that condition (1.6) and spatial homogeneity imply that $\mathbb{T}^*$ is a GW tree.
- All through this work, we shall assume that the P^*-BMC is super-critical.

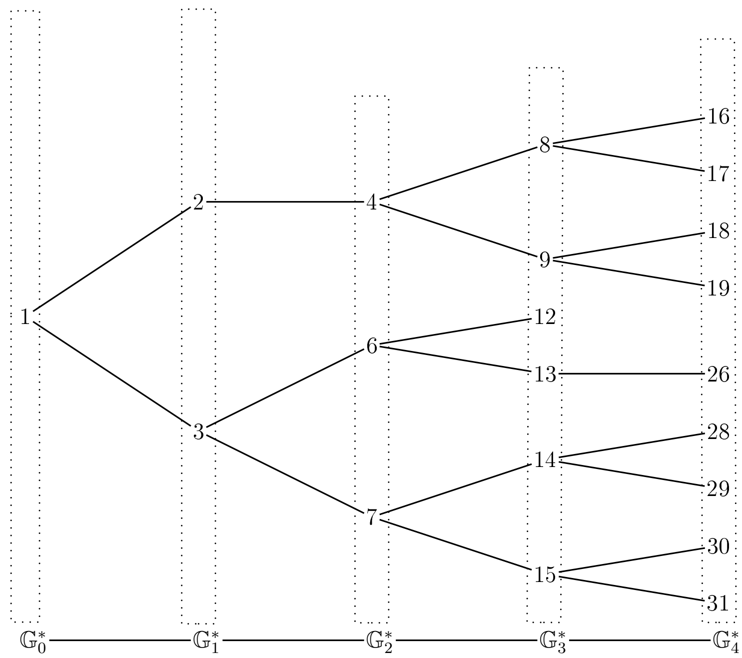

FIGURE 2. A binary GW tree up to the 4 th generation. *In this tree, individual 1 gives birth to two individuals which both divide, this happen with probability* $p_{1,0}$. *Individual 2 gives birth to two individuals which only one (the new pole) divides, this happen which probability* p_0. *Individual 12 gives birth to two individuals which do not divide, this happen with probability* $1 - p_{1,0} - p_0 - p_1$.

Now, for any subset $J \subset \mathbb{T}$, let

$$J^* = J \cap \mathbb{T}^* = \{j \in J : X_j \neq \partial\}$$

be the subset of living cells among J, and $|J|$ be the cardinal of J. The process $(|\mathbb{G}_k^*|, k \in \mathbb{N})$, is a GW process with the reproduction generating function

$$\psi(z) = (1 - p_0 - p_1 - p_{1,0}) + (p_0 + p_1)z + p_{1,0}z^2,$$

and the average number of daughters alive is m. It is known, see e.g [3], that $m^{-k}|\mathbb{G}_k^*|$ converges in probability to a non-negative random variable W. Moreover, $\mathbb{P}(W > 0) = 1$ iff there is no extinction. We have for all $r \geq 0$,

$$\mathbb{E}\left[|\mathbb{G}_r^*|\right] = m^r \quad \text{and} \quad \mathbb{E}\left[|\mathbb{T}_r^*|\right] = \sum_{q=0}^{r} \mathbb{E}\left[|\mathbb{G}_q^*|\right] = \frac{m^{r+1} - 1}{m - 1} := t_r. \tag{1.7}$$

It is known, see [12], that $t_r^{-1}|\mathbb{T}_r^*|$ converges in probability to W as well.
For $i \in \mathbb{T}$, set $\Delta_i = (X_i, X_{2i}, X_{2i+1})$ the mother-daughters quantities of interest. For a finite subset $J \subset \mathbb{T}$, we set

$$M_J(f) = \begin{cases} \sum_{i\in J} f(X_i) & \text{for } f \in \mathcal{B}(\bar{S}), \\ \sum_{i\in J} f(\Delta_i) & \text{for } f \in \mathcal{B}(\bar{S}^3), \end{cases} \tag{1.8}$$

with the convention that a sum over an empty set is null. We also define the following two averages of f over J

$$\overline{M}_J(f) = \frac{1}{|J|} M_J(f) \quad \text{if } |J| > 0 \quad \text{and} \quad \widetilde{M}_J(f) = \frac{1}{\mathbb{E}\left[|J|\right]} M_J(f) \quad \text{if } \mathbb{E}\left[|J|\right] > 0. \tag{1.9}$$

Limit theorems for averages (1.9) have been studied in [12] for $J = \mathbb{G}_n^*$ and $J = \mathbb{T}_n^*$, as n goes to infinity. Under uniform geometric ergodicity assumption for Q, we will establish in this paper deviation inequalities for those averages. These deviation inequalities will allow to highlight three regimes for the speed of convergence of above averages, thus showing a competition between the ergodicity of the embedded Markov chain $(Y_n, n \in \mathbb{N})$ and the size of the binary Galton-Watson tree. This new phenomenon is not observed in the asymptotic study of Delmas and Marsalle [12]. Notice that deviation inequalities were already studied in the no death case [8], that is $m = 2$. We will follow essentially the same approach that the latter paper for the proofs of our results. However, we will introduce some modifications on those proofs in order to take into account the randomness of index set, and we will make use of the theory of large deviation for branching processes [2]. Let us also mention [7], where the authors establish deviation inequalities for estimators of parameters of the p-order bifurcating autoregressive process.
The rest of paper is organized as follows. In section 2, we states our main results, that is deviation inequalities for averages (1.9), for $J = \mathbb{G}_n^*$ and $J = \mathbb{T}_n^*$. This will be done under uniform geometric ergodicity assumption for Q, and suitable assumptions on the binary GW tree. In section 3, we will focus in particular on the first order bifurcating autoregressive process with missing data described in section 1.1. Section 4 is dedicated to the proofs of our results.

2. Main results

We consider the following hypothesis:

- **(H1):** There exists a probability measure μ on $(S, \mathcal{S})$ such that for all $f \in \mathcal{B}_b(S)$ with $\langle \mu, f\rangle = 0$, there is $c > 0$ such that for all $k \in \mathbb{N}$ and for all $x \in S$, $|Q^k f(x)| \leq c\alpha^k$.
- **(H2):** $m > \sqrt{2}$.
- **(H3):** $p_{1,0} + p_0 + p_1 = 1$, where $p_{1,0}$, p_0 and p_1 are defined in section 1.1.

Remark 2.1. Hypothesis **(H1)** implies that the Markov chain Y is ergodic, that is for all $f \in \mathcal{C}_b(S)$ and for all $x \in S$, $\lim_{k\to\infty} \mathbb{E}_x[f(Y_k)] = \langle \mu, f\rangle$. Assuming hypothesis **(H3)** means that we work conditionally to the non-extinction. Note that this is consistent with the study of E. Coli.

Hypothesis (**H2**) comes from our calculations. indeed, in order to get relevant inequalities, i.e. inequalities for which the upper bound goes to zero as the sample size increases, we have

to assume that $m > \sqrt{2}$. However, our deviation inequalities also work for $m \leq \sqrt{2}$, but they are not relevant for this case. To get relevant deviation inequalities for $m \leq \sqrt{2}$ is still an open problem that we will pursue in an other work.

In the sequel, $\mathbb{H}_r$ will denote one of the set $\mathbb{G}_r$ or $\mathbb{T}_r$. We set $h_r = (m^2/2)^r$ if $\mathbb{H}_r = \mathbb{G}_r$ and $h_r = (m^2/2)^{r+1}$ if $\mathbb{H}_r = \mathbb{T}_r$. We can now state our main results. Notice that any function f defined on S is extended to $\bar{S}$ by setting $f(\partial) = 0$.

Theorem 2.2. *Under hypothesis* **(H1)** *and* **(H2)**, *let $f \in \mathcal{B}_b(S)$ such that $\langle \mu, f \rangle = 0$. Then we have for all $\delta > 0$:*

- *if $m\alpha < 1$, then $\forall r \in \mathbb{N}$,*
$$\mathbb{P}\Big(\widetilde{M}_{\mathbb{H}_r^*}(f) > \delta\Big) \leq \exp\left(c''\delta\right) \exp\left(-c'\delta^2 h_r\right);$$
- *if $m\alpha = 1$, then for $\mathbb{H}_r = \mathbb{G}_r$ and $\forall r \in \mathbb{N}$,*
$$\mathbb{P}\Big(\widetilde{M}_{\mathbb{H}_r^*}(f) > \delta\Big) \leq \exp\left(c''\delta\right) \exp\left(-c'\delta^2 (m^2/2)^r\right);$$
- *if $m\alpha = 1$, then for $\mathbb{H}_r = \mathbb{T}_r$ and $\forall r \in \mathbb{N}$,*
$$\mathbb{P}\Big(\widetilde{M}_{\mathbb{H}_r^*}(f) > \delta\Big) \leq \exp\left(c''\delta(r+1)\right) \exp\left(-c'\delta^2 (m^2/2)^{r+1}\right);$$
- *if $1 < m\alpha < \sqrt{2}$, then $\forall r \in \mathbb{N}$ such that $r > r_0$,*
$$\mathbb{P}\Big(\widetilde{M}_{\mathbb{H}_r^*}(f) > \delta\Big) \leq \exp\left(-c'\delta^2 h_r\right);$$
- *if $m\alpha = \sqrt{2}$, then $\forall r \in \mathbb{N}$ such that $r > r_0$,*
$$\mathbb{P}\Big(\widetilde{M}_{\mathbb{H}_r^*}(f) > \delta\Big) \leq \exp\left(-\frac{c'\delta^2 h_r}{r}\right);$$
- *if $m\alpha > \sqrt{2}$, then $\forall r \in \mathbb{N}^*$ such that $r > r_0$,*
$$\mathbb{P}\Big(\widetilde{M}_{\mathbb{H}_r^*}(f) > \delta\Big) \leq \exp\left(-\frac{c'\delta^2}{\alpha^{2r}}\right);$$

where,

- *$r_0 := \log\left(\frac{\delta}{c_0}\right) / \log(\alpha) - k_0$, with $k_0 \in \{0, 1\}$,*
- *c_0, c' and c'' are positive constants which depend on α, m, and c and may differ line by line.*

Theorem 2.3. *Under hypothesis* **(H1)**-**(H3)**, *we have for all $f \in \mathcal{B}_b(S)$ such that $\langle \mu, f \rangle \neq 0$ and for all $\delta > 0$:*

- *if $m\alpha < 1$, then $\forall r \in \mathbb{N}$,*
$$\mathbb{P}\Big(\widetilde{M}_{\mathbb{H}_r^*}(f) - \langle \mu, f \rangle W > \delta\Big) \leq \exp\left(c''\delta\right) \exp\left(-c'\delta^2 h_r\right) + A_r;$$
- *if $m\alpha = 1$, then for $\mathbb{H}_r = \mathbb{G}_r$ and $\forall r \in \mathbb{N}$,*
$$\mathbb{P}\Big(\widetilde{M}_{\mathbb{H}_r^*}(f) - \langle \mu, f \rangle W > \delta\Big) \leq \exp\left(c''\delta\right) \exp\left(-c'\delta^2 (m^2/2)^r\right) + A_r;$$

- *if* $m\alpha = 1$, *then for* $\mathbb{H}_r = \mathbb{T}_r$ *and* $\forall r \in \mathbb{N}$,
$$\mathbb{P}\Big(\widetilde{M}_{\mathbb{H}_r^*}(f) - \langle \mu, f\rangle W > \delta\Big) \leq \exp\left(c''\delta(r+1)\right)\exp\left(-c'\delta^2(m^2/2)^{r+1}\right) + A_r;$$
- *if* $1 < m\alpha < \sqrt{2}$, *then* $\forall r \in \mathbb{N}$ *such that* $r > r_0$,
$$\mathbb{P}\Big(\widetilde{M}_{\mathbb{H}_r^*}(f) - \langle \mu, f\rangle W > \delta\Big) \leq \exp\left(-c'\delta^2 h_r\right) + A_r;$$
- *if* $m\alpha = \sqrt{2}$, *then* $\forall r \in \mathbb{N}$ *such that* $r > r_0$,
$$\mathbb{P}\Big(\widetilde{M}_{\mathbb{H}_r^*}(f) - \langle \mu, f\rangle W > \delta\Big) \leq \exp\left(-\frac{c'\delta^2 h_r}{r}\right) + A_r;$$
- *if* $m\alpha > \sqrt{2}$, *then* $\forall r \in \mathbb{N}^*$ *such that* $r > r_0$,
$$\mathbb{P}\Big(\widetilde{M}_{\mathbb{H}_r^*}(f) - \langle \mu, f\rangle W > \delta\Big) \leq \exp\left(-\frac{c'\delta^2}{\alpha^{2r}}\right) + A_r;$$

where,

- *for all* $r \in \mathbb{N}$,
$$A_r = \begin{cases} c'\exp\left(-c''\delta^{2/3}(m^{1/3})^r\right) & \text{if } \mathbb{H}_r = \mathbb{G}_r \\ \exp\left(c'\delta^{2/3}\right)\exp\left(-c''\delta^{2/3}\left(t_r/(r+1)^2\right)^{1/3}\right) & \text{if } \mathbb{H}_r = \mathbb{T}_r, \end{cases}$$
- $r_0 := \log\left(\frac{\delta}{c_0}\right)/\log(\alpha) - k_0$, *with* $k_0 \in \{0, 1\}$,
- c_0, c' *and* c'' *are positive constants which depend on* α, m, *and* c *and may differ line by line.*

Remark 2.4. For $\langle \mu, f\rangle = 0$ (Theorem 2.2), there is no additional term in the deviation of average $\widetilde{M}_{\mathbb{H}_r^*}(f)$. While in Theorem 2.3 there is an additional term A_r which appears. This term is a contribution of the binary Galton-Watson tree on the deviation of average $\widetilde{M}_{\mathbb{H}_r^*}(f)$ with respect to $\langle \mu, f\rangle W$. This explain why we need additional hypothesis (**H3**) in Theorem 2.3, because we have to deal with the deviation inequalities for Galton-Watson processes.

The next results can be seen as a consequence of the previous results.

Theorem 2.5. *We assume that hypothesis* **(H1)-(H3)** *are satisfied. Let* $f \in \mathcal{B}_b(S)$. *For all* $\delta > 0$, *for all* $a > 0$ *and for all* $b > 0$ *such that* $b < a/(\delta+1)$, *we have*

- *if* $m\alpha < 1$, *then* $\forall r \in \mathbb{N}$,
$$\mathbb{P}\Big(\overline{M}_{\mathbb{H}_r^*}(f) - \langle \mu, f\rangle > \delta \Big| W \geq a\Big) \leq \exp\left(c''\delta b\right)\exp\left(-c'(\delta b)^2 h_r\right) + A_r;$$
- *if* $m\alpha = 1$, *then for* $\mathbb{H}_r = \mathbb{G}_r$ *and* $\forall r \in \mathbb{N}$,
$$\mathbb{P}\Big(\overline{M}_{\mathbb{H}_r^*}(f) - \langle \mu, f\rangle > \delta \Big| W \geq a\Big) \leq \exp\left(c''\delta b\right)\exp\left(-c'(\delta b)^2(m^2/2)^r\right) + A_r;$$
- *if* $m\alpha = 1$, *then for* $\mathbb{H}_r = \mathbb{T}_r$ *and* $\forall r \in \mathbb{N}$,
$$\mathbb{P}\Big(\overline{M}_{\mathbb{H}_r^*}(f) - \langle \mu, f\rangle > \delta \Big| W \geq a\Big) \leq \exp\left(c''\delta b(r+1)\right)\exp\left(-c'(\delta b)^2(m^2/2)^{r+1}\right) + A_r;$$
- *if* $1 < m\alpha < \sqrt{2}$, *then* $\forall r \in \mathbb{N}$ *such that* $r > r_0$,
$$\mathbb{P}\Big(\overline{M}_{\mathbb{H}_r^*}(f) - \langle \mu, f\rangle > \delta \Big| W \geq a\Big) \leq \exp\left(-c'(\delta b)^2 h_r\right) + A_r;$$

- *if* $m\alpha = \sqrt{2}$, *then* $\forall r \in \mathbb{N}$ *such that* $r > r_0$,

$$\mathbb{P}\Big(\overline{M}_{\mathbb{H}_r^*}(f) - \langle \mu, f\rangle > \delta \Big| W \geq a\Big) \leq \exp\left(-\frac{c'(\delta b)^2 h_r}{r}\right) + A_r;$$

- *if* $m\alpha > \sqrt{2}$, *then* $\forall r \in \mathbb{N}^*$ *such that* $r > r_0$,

$$\mathbb{P}\Big(\overline{M}_{\mathbb{H}_r^*}(f) - \langle \mu, f\rangle > \delta \Big| W \geq a\Big) \leq \exp\left(-\frac{c'(\delta b)^2}{\alpha^{2r}}\right) + A_r;$$

where,

- *for all* $r \in \mathbb{N}$,

$$A_r = \begin{cases} c' \exp\left(-c''(\delta b)^{2/3}(m^{1/3})^r\right) & \text{if } \mathbb{H}_r = \mathbb{G}_r \\ \exp\left(c'(\delta b)^{2/3}\right)\exp\left(-c''(\delta b)^{2/3}\left(t_r/(r+1)^2\right)^{1/3}\right) & \text{if } \mathbb{H}_r = \mathbb{T}_r, \end{cases}$$

- $r_0 := \log\left(\frac{\delta b}{c_0}\right)/\log(\alpha) - k_0$, *with* $k_0 \in \{0, 1\}$,
- c_0, c' *and* c'' *are positive constants which depend on* α, m, a, *and* c, *and may differ line by line.*

We have the following extension of above theorems when f does not only depend on an individual X_i, but on the mother-daughters triangle Δ_i.

Theorem 2.6. *Let* $f \in \mathcal{B}_b(S^3)$. *If* $\langle \mu, P^* f\rangle = 0$, *then, under hypothesis* ***(H1)*** *and* ***(H2)***, *we have deviation inequalities of Theorem 2.2 for* $\widetilde{M}_{\mathbb{H}_r^*}(f)$. *If* $\langle \mu, P^* f\rangle \neq 0$, *under additional hypothesis* ***(H3)***, *we have deviation inequalities of Theorem 2.3 for* $\widetilde{M}_{\mathbb{H}_r^*}(f) - \langle \mu, P^* f\rangle W$ *and of Theorem 2.5 for* $\overline{M}_{\mathbb{H}_r^*}(f) - \langle \mu, P^* f\rangle$.

Remark 2.7. Let us stress that by tedious, but straightforward calculations, the constants which appear in the previous inequalities can be made explicit.

Let us recall the following definition.

Definition 2.8. *Let* (E, d) *be a metric space. Let* (Z_n) *be a sequence of random variables valued in* E, Z *be a random variable valued in* E *and* (v_n) *be a rate. We say that* Z_n *converges* v_n*-superexponentially fast in probability to* Z *if for all* $\delta > 0$,

$$\limsup_{n\to\infty} \frac{1}{v_n} \log \mathbb{P}(d(Z_n, Z) > \delta) = -\infty.$$

This "exponential convergence" with speed v_n *will be shortened as*

$$Z_n \overset{\text{superexp}}{\underset{v_n}{\longrightarrow}} Z.$$

Remark 2.9. Let (b_n) be a sequence of increasing positive real numbers such that

$$b_n \to +\infty$$

and

- if $m\alpha < \sqrt{2}$, the sequence (b_n) is such that $b_n/\sqrt{n} \longrightarrow 0$,
- if $m\alpha = \sqrt{2}$, the sequence (b_n) is such that $(b_n\sqrt{\log n})/\sqrt{n} \longrightarrow 0$,
- if $m\alpha > \sqrt{2}$, the sequence (b_n) is such that $b_n \alpha^{\log n/\log(m^2/2)} \longrightarrow 0$.

From the previous deviations inequalities, we can deduce easily that

$$\widetilde{M}_{\mathbb{H}_r^*}(f) \underset{b^2_{\lfloor h_r \rfloor}}{\overset{\text{superexp}}{\longrightarrow}} 0 \quad \text{if } \langle \mu, f \rangle = 0,$$

and if $\langle \mu, f \rangle \neq 0$, we have for $m < 2^{3/5}$

$$\widetilde{M}_{\mathbb{H}_r^*}(f) \underset{b^2_{\lfloor h_r \rfloor}}{\overset{\text{superexp}}{\longrightarrow}} \langle \mu, f \rangle W,$$

and $\forall a > 0$,

$$\limsup_{r \to +\infty} \frac{1}{b^2_{\lfloor h_r \rfloor}} \log \mathbb{P}\left(\left| \overline{M}_{\mathbb{H}_r^*}(f) - \langle \mu, f \rangle \right| > \delta \middle| W \geq a \right) = -\infty.$$

So, for the exponential convergence of averages $\widetilde{M}_{\mathbb{H}_r^*}(f)$ and $\overline{M}_{\mathbb{H}_r^*}(f)$, there are three regimes according to the value of $m\alpha$ compared to $\sqrt{2}$. This phenomenon is not observed in the limit theorems of Delmas and Marsalle [12]. However, a similar phenomenon was observed recently by Adamczak and Miłoś for the central limit theorem of branching particle system [1].

So, our deviations inequalities highlight a competition between the ergodicity of the embedded Markov chain with transition probability Q and the Galton-Watson binary tree.

3. Application:First order bifurcating autoregressive processes with missing data

We consider the asymmetric auto-regressive processes given in section 1.1. Notice that the process $(X_i, i \in \mathbb{T})$ defined in section 1.1, with the convention that $X_i = \partial$ if the cell i is missing, is a spatially homogeneous BMC on a GW tree. We will assume that $2p_{1,0}+p_1+p_0 > \sqrt{2}$. This implies in particular that the BMC on GW is super-critical. We will also assume that the noise sequences $((\varepsilon_{2n}, \varepsilon_{2n+1}), n \in \mathbb{T})$, $(\varepsilon'_{2n}, n \in \mathbb{T})$ and $(\varepsilon'_{2n+1}, n \in \mathbb{T})$, and the initial state X_1 take their values in a compact set. The latter implies that the process $(X_i, i \in \mathbb{T})$ is bounded. We denote by S the state space of $(X_i, i \in \mathbb{T})$. We assume without loss of generality that S is a compact subset of $\mathbb{R}$.

Let $\mathbb{T}_n^{0,1}$ be the subset of cells in $\mathbb{T}_n^*$ with two living daughters, $\mathbb{T}_n^0$ (resp.$\mathbb{T}_n^1$) be the set of cells of $\mathbb{T}_n^*$ with only the new (resp. old) pole daughter alive:

$$\mathbb{T}_n^{1,0} = \left\{ i \in \mathbb{T}_n^* : \Delta_i \in S^3 \right\}, \qquad \mathbb{T}_n^0 = \{ i \in \mathbb{T}_n^* : \Delta_i \in S^2 \times \{\partial\}\} \qquad \text{and}$$
$$\mathbb{T}_n^1 = \{ i \in \mathbb{T}_n^* : \Delta_i \in S \times \{\partial\} \times S \}.$$

We compute the least-squares estimator (LSE)

$$\widehat{\theta}_n = (\widehat{\alpha}_0^n, \widehat{\beta}_0^n, \widehat{\alpha}_1^n, \widehat{\beta}_1^n, \widehat{\alpha}_0'^n, \widehat{\beta}_0'^n, \widehat{\alpha}_1'^n, \widehat{\beta}_1'^n)$$

of θ given by (1.5), based on the observation of a sub-tree $\mathbb{T}_{n+1}^*$. Consequently, we obviously have for $\eta \in \{0, 1\}$,

$$\widehat{\alpha}_\eta^n = \frac{|\mathbb{T}_n^{1,0}|^{-1} \sum_{i \in \mathbb{T}_n^{1,0}} X_i X_{2i+\eta} - \left(|\mathbb{T}_n^{1,0}|^{-1} \sum_{i \in \mathbb{T}_n^{1,0}} X_i \right) \left(|\mathbb{T}_n^{1,0}|^{-1} \sum_{i \in \mathbb{T}_n^{1,0}} X_{2i+\eta} \right)}{|\mathbb{T}_n^{1,0}|^{-1} \sum_{i \in \mathbb{T}_n^{1,0}} X_i^2 - \left(|\mathbb{T}_n^{1,0}|^{-1} \sum_{i \in \mathbb{T}_n^{1,0}} X_i \right)^2},$$

$$\widehat{\beta}_\eta^n = |\mathbb{T}_n^{1,0}|^{-1} \sum_{i\in\mathbb{T}_n^{1,0}} X_{2i+\eta} - \widehat{\alpha}_\eta^n |\mathbb{T}_n^{1,0}|^{-1} \sum_{i\in\mathbb{T}_n^{1,0}} X_i,$$

$$\widehat{\alpha}_\eta'^n = \frac{|\mathbb{T}_n^\eta|^{-1} \sum_{i\in\mathbb{T}_n^\eta} X_i X_{2i+\eta} - \left(|\mathbb{T}_n^\eta|^{-1} \sum_{i\in\mathbb{T}_n^\eta} X_i\right)\left(|\mathbb{T}_n^\eta|^{-1} \sum_{i\in\mathbb{T}_n^\eta} X_{2i+\eta}\right)}{|\mathbb{T}_n^\eta|^{-1} \sum_{i\in\mathbb{T}_n^\eta} X_i^2 - \left(|\mathbb{T}_n^\eta|^{-1} \sum_{i\in\mathbb{T}_n^\eta} X_i\right)^2},$$

$$\widehat{\beta}_\eta'^n = |\mathbb{T}_n^\eta|^{-1} \sum_{i\in\mathbb{T}_n^\eta} X_{2i+\eta} - \widehat{\alpha}_\eta'^n |\mathbb{T}_n^\eta|^{-1} \sum_{i\in\mathbb{T}_n^\eta} X_i.$$

Notice that those LSE are based on polynomial functions of the observations. So, since the latter are bounded, we are in the functional setting of the results of section 2. Recalling the Markov chain $(Y_n, n \in \mathbb{N})$, notice that Y_n is distributed as $Z_n = a_1a_2\cdots a_{n-1}a_nY_0 + \sum_{k=1}^n a_1a_2\cdots a_{k-1}b_k$, where $b_n = b'_n + s_ne_n$, $((a_n, b'_n, s_n), n \geq 1)$ is a sequence of independent identically distributed random variables, whose common distribution is given by, for $\eta \in \{0,1\}$,

$$\mathbb{P}(a_1 = \alpha_\eta, b'_1 = \beta_\eta, s_1 = \sigma) = \frac{p_{1,0}}{m} \quad \text{and} \quad \mathbb{P}(a_1 = \alpha'_\eta, b'_1 = \beta'_\eta, s_1 = \sigma_\eta) = \frac{p_\eta}{m},$$

$(e_n, n \geq 1)$ is a sequence of independent $\mathcal{N}(0,1)$ random variables, and is independent of $((a_n, b'_n, s_n), n \geq 1)$, and both sequences are independent of Y_0. Moreover, it is easy to check that the sequence $(Z_n, n \in \mathbb{N})$ converge a.s. to a limit Z, which implies that the Markov chain $(Y_n, n \in \mathbb{N})$ converge in distribution to Z. We refer to [12], section 6, for more details. Following the proof of Proposition 28, step 1 in [14], we check hypothesis **(H1)** with $\alpha = \max(|\alpha_0|, |\alpha_1|, |\alpha'_0|, |\alpha'_1|) < 1$ and with μ the distribution of Z. Let $\mu_1 = \mathbb{E}[Z]$ and $\mu_2 = \mathbb{E}\left[Z^2\right]$. We have (see [12])

$$\mu_1 = \frac{\overline{\beta}}{1-\overline{\alpha}} \quad \text{and} \quad \mu_2 = \frac{2\overline{\alpha\beta}\overline{\beta}/(1-\overline{\alpha}) + \overline{\beta^2} + \overline{\alpha^2}}{1-\overline{\alpha^2}},$$

where $\overline{\alpha} = \mathbb{E}[a_1]$, $\overline{\alpha^2} = \mathbb{E}\left[a_1^2\right]$, $\overline{\beta} = \mathbb{E}[b_1]$, $\overline{\beta^2} = \mathbb{E}\left[b_1^2\right]$, $\overline{\alpha\beta} = \mathbb{E}[a_1b_1]$ and $\overline{\sigma^2} = \mathbb{E}\left[s_1^2\right]$.
We then have the following deviation inequality for $\widehat{\theta}_n - \theta$.

Proposition 3.1. *For all $\delta > 0$, for all $a > 0$, for all $b > 0$ and for all $\gamma > 0$ such that $b < a/(\delta+1)$ and $\gamma < \min\left\{c_1/(1+\delta), c_1/\left(1+\sqrt{\delta}\right)\right\}$, where c_1 is a positive constant which depends on $p_{1,0}$, p_0, p_1, μ_1 and μ_2, and for $n_0 := (\log(\gamma^q\delta^p b/c_0)/\log\alpha) - 1$, we have*

- *if $m\alpha < 1$, then $\forall n \in \mathbb{N}$,*
$$\mathbb{P}\left(\|\widehat{\theta}_n - \theta\| > \delta | W \geq a\right) \leq c_2 \exp\left(c''\gamma^q\delta^p b\right) \exp\left(-c'\left(\gamma^q\delta^p b\right)^2 \left(m^2/2\right)^{n+1}\right) + A_n;$$
- *if $m\alpha = 1$, then $\forall n \in \mathbb{N}$,*
$$\mathbb{P}\left(\|\widehat{\theta}_n - \theta\| > \delta | W \geq a\right) \leq c_2 \exp\left(c''\gamma^q\delta^p b(n+1)\right) \exp\left(-c'\left(\gamma^q\delta^p b\right)^2 \left(m^2/2\right)^{n+1}\right) + A_n;$$
- *if $1 < m\alpha < \sqrt{2}$, then $\forall n \in \mathbb{N}$ such that $n > n_0$,*
$$\mathbb{P}\left(\|\widehat{\theta}_n - \theta\| > \delta | W \geq a\right) \leq c_2 \exp\left(-c'\left(\gamma^q\delta^p b\right)^2 \left(m^2/2\right)^{n+1}\right) + A_n;$$

- *if* $m\alpha = \sqrt{2}$, *then* $\forall n \in \mathbb{N}$ *such that* $n > n_0$,
$$\mathbb{P}\left(\|\widehat{\theta}_n - \theta\| > \delta | W \geq a\right) \leq c_2 \exp\left(-c' \left(\gamma^q \delta^p b\right)^2 (1/n) \left(m^2/2\right)^{n+1}\right) + A_n;$$
- *if* $m\alpha > \sqrt{2}$, *then* $\forall n \in \mathbb{N}^*$ *such that* $n > n_0$,
$$\mathbb{P}\left(\|\widehat{\theta}_n - \theta\| > \delta | W \geq a\right) \leq c_2 \exp\left(-c' \left(\gamma^q \delta^p b\right)^2 \alpha^{-2n}\right) + A_n;$$

where $A_n = c_3 \exp\left(c' \left(\gamma^q \delta^p b\right)^{2/3}\right) \exp\left(-c'' \left(\gamma^q \delta^p b\right)^{2/3} \left(t_n/(n+1)^2\right)^{1/3}\right)$, $p \in \{1/2, 1\}$, $q \in \{0, 1/2, 1\}$, c_2, c_3, c_4, c' *and* c'' *are positive constants which depend on* c, m, α $p_{1,0}$, p_0, p_1, μ_1 *and* μ_2.

Remark 3.2. Note that the constants c_2, c_3, c_4, c' and c'' which appear in Proposition 3.1 may differ term by term. The values of p and q depend on the magnitude of δ and γ. For example, for δ and γ small enough, we have $p = 1$ and $q = 1$. We also stress that all these constants can be made explicit by tedious calculations.

4. Proofs of the main results

4.1. **Proof of Theorem 2.2.** Let $f \in \mathcal{B}_b(S)$ such that $\langle \mu, f \rangle = 0$. We are going to study successively $\widetilde{M}_{\mathbb{H}_r^*}(f)$ for $\mathbb{H}_r = \mathbb{G}_r$ and $\mathbb{H}_r = \mathbb{T}_r$.

Step 1. Let us first deal with $\widetilde{M}_{\mathbb{G}_r^*}(f)$. By Chernoff inequality, we have for all $\delta > 0$ and for all $\lambda > 0$

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) > \delta\right) \leq \exp\left(-\lambda \delta m^r\right) \mathbb{E}\left[\exp\left(\lambda \sum_{i \in \mathbb{G}_r^*} f(X_i)\right)\right]. \tag{4.1}$$

Recall that for all $i \in \mathbb{G}_{r-1}^*$,

$$\mathbb{E}\left[f(X_{2i}) \mathbf{1}_{\{2i \in \mathbb{T}^*\}} + f(X_{2i+1}) \mathbf{1}_{\{2i+1 \in \mathbb{T}^*\}} | \mathcal{F}_{r-1}\right] = mQf(X_i).$$

By subtracting and adding terms in expectation of the right hand of (4.1), and conditioning with respect to $\mathcal{F}_{r-1}$, we get

$$\begin{aligned} \mathbb{E}\left[\exp\left(\lambda \sum_{i \in \mathbb{G}_r^*} f(X_i)\right)\right] &= \mathbb{E}\left[\exp\left(\lambda \sum_{i \in \mathbb{G}_{r-1}^*} mQf(X_i)\right)\right. \\ &\times \mathbb{E}\left[\exp\left(\sum_{i \in \mathbb{G}_{r-1}^*} \lambda \left(f(X_{2i}) \mathbf{1}_{\{2i \in \mathbb{T}^*\}} + f(X_{2i+1}) \mathbf{1}_{\{2i+1 \in \mathbb{T}^*\}} - mQf(X_i)\right)\right) \Bigg| \mathcal{F}_{r-1}\right]\Bigg]. \end{aligned} \tag{4.2}$$

Observing that $\mathbb{G}_{r-1}^*$ is $\mathcal{F}_{r-1}$ measurable, and using the fact that conditionally to $\mathcal{F}_{r-1}$, the triplets $\{(\Delta_i), i \in \mathbb{G}_{r-1}\}$ are independent (this is due to the Markov property), we have

$$\begin{aligned} &\mathbb{E}\left[\exp\left(\sum_{i \in \mathbb{G}_{r-1}^*} \lambda \left(f(X_{2i}) \mathbf{1}_{\{2i \in \mathbb{T}^*\}} + f(X_{2i+1}) \mathbf{1}_{\{2i+1 \in \mathbb{T}^*\}} - mQf(X_i)\right)\right) \Bigg| \mathcal{F}_{r-1}\right] \\ &= \prod_{i \in \mathbb{G}_{r-1}^*} \mathbb{E}\left[\exp\left(\lambda \left(f(X_{2i}) \mathbf{1}_{\{2i \in \mathbb{T}^*\}} + f(X_{2i+1}) \mathbf{1}_{\{2i+1 \in \mathbb{T}^*\}} - mQf(X_i)\right)\right) \Bigg| \mathcal{F}_{r-1}\right]. \end{aligned} \tag{4.3}$$

Using Azuma-Bennet-Hoeffding inequality [4], [5], [16], we get according to **(H1)**, for all $i \in \mathbb{G}^*_{r-1}$,

$$\mathbb{E}\left[\exp\left(\lambda\left(f(X_{2i})\mathbf{1}_{\{2i\in\mathbb{T}^*\}} + f(X_{2i+1})\mathbf{1}_{\{2i+1\in\mathbb{T}^*\}} - mQf(X_i)\right)\right)\middle|\mathcal{F}_{r-1}\right] \leq \exp\left(\frac{c^2\lambda^2(2+m\alpha)^2}{2}\right).$$

From (4.3), this implies that

$$\mathbb{E}\left[\exp\left(\sum_{i\in\mathbb{G}^*_{r-1}}\lambda\left(f(X_{2i})\mathbf{1}_{\{2i\in\mathbb{T}^*\}} + f(X_{2i+1})\mathbf{1}_{\{2i+1\in\mathbb{T}^*\}} - mQf(X_i)\right)\right)\middle|\mathcal{F}_{r-1}\right] \leq \exp\left(\frac{c^2\lambda^2(2+m\alpha)^2|\mathbb{G}^*_{r-1}|}{2}\right) \leq \exp\left(\frac{c^2\lambda^2(2+m\alpha)^2|\mathbb{G}_{r-1}|}{2}\right),$$

where we have used the fact that $|\mathbb{G}^*_{r-1}| \leq |\mathbb{G}_{r-1}|$ in the last inequality. Recalling (4.2), we are led to

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}^*_r} f(X_i)\right)\right] \leq \exp\left(\frac{c^2\lambda^2(2+m\alpha)^2|\mathbb{G}_{r-1}|}{2}\right) \times\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}^*_{r-1}} mQf(X_i)\right)\right].$$

Reproducing the same reasoning with Qf and $\mathbb{G}^*_{r-1}$ instead of f and $\mathbb{G}^*_r$, we get

$$\mathbb{E}\left[\exp\left(\lambda m\sum_{i\in\mathbb{G}^*_{r-1}} Qf(X_i)\right)\right] \leq \exp\left(\frac{c^2\lambda^2m^2(2\alpha+m\alpha^2)^2|\mathbb{G}_{r-2}|}{2}\right) \times\mathbb{E}\left[\exp\left(\lambda m^2\sum_{i\in\mathbb{G}^*_{r-2}} Q^2f(X_i)\right)\right].$$

Iterating this procedure, we get

$$\begin{aligned}\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}^*_r} f(X_i)\right)\right] &\leq \exp\left(\frac{c^2\lambda^2}{2}\sum_{q=0}^{r-1}\left(2\alpha^q + m\alpha^{q+1}\right)^2 m^{2q}2^{r-1-q}\right) \times \mathbb{E}\Big[\exp\left(\lambda m^r Q^r f(X_1)\right)\Big] \\ &\leq \exp\left(\frac{c^2\lambda^2(2+m\alpha)^2 2^{r-1}}{2}\sum_{q=0}^{r-1}\left(\frac{\alpha^2m^2}{2}\right)^q\right) \times \exp\left(\lambda c(\alpha m)^r\right),\end{aligned}$$

where the last inequality was obtained from **(H1)**. From the foregoing and from (4.1), we deduce that

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) > \delta\right) \leq \begin{cases} \exp\left(-\lambda\delta m^r + \frac{c^2\lambda^2(2+m\alpha)^2\left(2^r-(\alpha^2m^2)^r\right)}{2(2-\alpha^2m^2)}\right) \\ \qquad\qquad\qquad \times \exp\left(\lambda c(\alpha m)^r\right) \quad \text{if } \alpha^2m^2 \neq 2, \\ \exp\left(-\lambda\delta m^r + c^2\lambda^2(2+\sqrt{2})^2 r 2^{r-2}\right)\exp\left(\lambda c(\sqrt{2})^r\right) \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{if } \alpha^2m^2 = 2. \end{cases}$$

Now, the rest divides into four cases. In the sequel c_1 and c_2 will denote positive constants which depend on c, m, and α.

• If $m\alpha \leq 1$, then, for all $r \in \mathbb{N}$, $(m\alpha)^r < 1$ and $2^r - (\alpha^2m^2)^r < 2^r$. We then have

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) > \delta\right) \leq \exp\left(c\lambda\right)\exp\left(-\lambda\delta m^r + \lambda^2 c_1 2^r\right).$$

Taking $\lambda = (\delta m^r)/(2^{r+1}c_1)$, we are led to

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) > \delta\right) \leq \exp\left(c_1\delta\right)\exp\left(-\delta^2 c_1\left(\frac{m^2}{2}\right)^r\right).$$

• If $1 < m\alpha < \sqrt{2}$, then, since $2^r - (\alpha^2m^2)^r < 2^r$, we have

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) > \delta\right) \leq \exp\left(-\lambda\delta m^r + \lambda^2 c_1 2^r\right)\exp\left(\lambda c(m\alpha)^r\right).$$

Taking $\lambda = (\delta m^r)/(2^{r+1}c_1)$, we are led to

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) > \delta\right) \leq \exp\left(-c_2\delta(m^2/2)^r(\delta - 2c\alpha^r)\right).$$

For all $r \in \mathbb{N}$ such that $r > \log(\delta/4c)/\log(\alpha)$, we have $\delta - 2c\alpha^r > \delta/2$ and it then follows that

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) > \delta\right) \leq \exp\left(-c_2\delta^2(m^2/2)^r\right).$$

• If $m\alpha = \sqrt{2}$, then we have

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) > \delta\right) \leq \exp\left(-\lambda\delta m^r + \lambda^2 c_1 r 2^{r-2})\right)\exp\left(\lambda c\left(\sqrt{2}\right)^r\right).$$

Taking $\lambda = (\delta m^r)/(c_1 r 2^{r-1})$, we have for all $r > \log(\delta/4c)/\log(\sqrt{2}/m)$,

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) > \delta\right) \leq \exp\left(-c_2\delta^2(1/r)(m^2/2)^r\right).$$

• If $m\alpha > \sqrt{2}$, then we have

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) > \delta\right) \leq \exp\left(-\lambda\delta m^r + \lambda^2 c_1(m^2\alpha^2)^r\right)\exp\left(\lambda c(m\alpha)^r\right).$$

Taking $\lambda = \delta/(2c_1(m\alpha^2)^r)$, we have for all $r > \log(\delta/4c)/\log\alpha$,

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) > \delta\right) \leq \exp\left(-c_3\delta^2\alpha^{-2r}\right).$$

This ends the proof for $\mathbb{H}_r = \mathbb{G}_r$.

Step 2. Let us look at $\widetilde{M}_{\mathbb{T}_r^*}(f)$. By Chernoff inequality, we have for all $\delta > 0$ and for all $\lambda > 0$

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \leq \exp\left(-\lambda\delta t_r\right)\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_r^*} f(X_i)\right)\right]. \tag{4.4}$$

Expectation which appears in the right hand of (4.4) can be written as

$$\begin{aligned}\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_r^*} f(X_i)\right)\right] = \mathbb{E}\Bigg[\exp\left(\lambda\sum_{i\in\mathbb{T}_{r-2}^*} f(X_i)\right) \\ \times\exp\left(\lambda\sum_{i\in\mathbb{G}_{r-1}^*}(f+mQf)(X_i)\right) \\ \times\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}_{r-1}^*}\left(f(X_{2i})\mathbf{1}_{\{2i\in\mathbb{T}^*\}} + f(X_{2i+1})\mathbf{1}_{\{2i+1\in\mathbb{T}^*\}} - mQf(X_i)\right)\right)\Big|\mathcal{F}_{r-1}\right]\Bigg].\end{aligned} \tag{4.5}$$

Observing that $\mathbb{G}_{r-1}^*$ is $\mathcal{F}_{r-1}$ measurable, and using the fact that conditionally to $\mathcal{F}_{r-1}$, the triplets $\{(\Delta_i), i\in\mathbb{G}_{r-1}\}$ are independent and Azuma-Bennet-Hoeffding inequality, we obtain

$$\begin{aligned}&\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}_{r-1}^*}\left(f(X_{2i})\mathbf{1}_{\{2i\in\mathbb{T}^*\}} + f(X_{2i+1})\mathbf{1}_{\{2i+1\in\mathbb{T}^*\}} - mQf(X_i)\right)\right)\Big|\mathcal{F}_{r-1}\right] \\ &\quad= \prod_{i\mathbb{G}_{r-1}^*}\mathbb{E}\left[\exp\left(\lambda\left(f(X_{2i})\mathbf{1}_{\{2i\in\mathbb{T}^*\}} + f(X_{2i+1})\mathbf{1}_{\{2i+1\in\mathbb{T}^*\}} - mQf(X_i)\right)\right)\Big|\mathcal{F}_{r-1}\right] \\ &\qquad\leq \exp\left(\frac{c^2\lambda^2(2+m\alpha)^2|\mathbb{G}_{r-1}^*|}{2}\right) \\ &\qquad\leq \exp\left(\frac{c^2(2+m\alpha)^2|\mathbb{G}_{r-1}|}{2}\right),\end{aligned}$$

where the last inequality was obtained using the fact that $|\mathbb{G}_{r-1}^*| \leq |\mathbb{G}_{r-1}|$. From the foregoing and from (4.5), we deduce that

$$\begin{aligned}\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_r^*} f(X_i)\right)\right] &\leq \exp\left(\frac{c^2(2+m\alpha)^2|\mathbb{G}_{r-1}|}{2}\right) \\ &\quad\times\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_{r-2}^*} f(X_i)\right)\exp\left(\lambda\sum_{i\in\mathbb{G}_{r-1}^*}(f+mQf)(X_i)\right)\right]\end{aligned}$$

Doing the same thing with $(f+mQf)$ and $\mathbb{G}_{r-1}^*$ instead of f and $\mathbb{G}_r^*$, we get

$$
\begin{aligned}
&\mathbb{E}\left[\exp\left(\lambda \sum_{i\in\mathbb{T}_{r-2}^*} f(X_i)\right)\exp\left(\lambda\sum_{i\in\mathbb{G}_{r-1}^*}(f+mQf)(X_i)\right)\right]\\
&\qquad = \mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_{r-3}^*} f(X_i)\right)\times\exp\left(\lambda\sum_{i\in\mathbb{G}_{r-2}^*}(f+mQf+m^2Q^2f)(X_i)\right)\right.\\
&\qquad\quad \times\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}_{r-2}^*}\Big((f+mQf)(X_{2i})\mathbf{1}_{\{2i\in\mathbb{T}^*\}}+(f+mQf)(X_{2i+1})\mathbf{1}_{\{2i+1\in\mathbb{T}^*\}}\right.\right.\\
&\qquad\qquad\qquad\qquad\qquad\qquad\left.\left.\left.-(mQf+m^2Q^2f)(X_i)\Big)\right)\Big|\mathcal{F}_{r-1}\right]\right]\\
&\qquad \le \exp\left(\frac{c^2\lambda^2(2+3m\alpha+m^2\alpha^2)^2|\mathbb{G}_{r-2}|}{2}\right)\\
&\qquad\quad\times\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_{r-3}^*} f(X_i)\right)\times\exp\left(\lambda\sum_{i\in\mathbb{G}_{r-2}^*}(f+mQf+m^2Q^2f)(X_i)\right)\right].
\end{aligned}
$$

Iterating this procedure, we are led to

$$
\begin{aligned}
\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_r^*} f(X_i)\right)\right] &\le \exp\left(\frac{c^2(2+m\alpha)^2\lambda^2}{2}\sum_{q=1}^{r}\left(\sum_{k=0}^{q-1}(m\alpha)^k\right)^2 2^{r-q}\right)\\
&\qquad\times\mathbb{E}\left[\exp\left(\lambda\sum_{q=0}^{r} m^qQ^qf(X_1)\right)\right]\\
&\le \exp\left(\frac{c^2(2+m\alpha)^2\lambda^2}{2}\sum_{q=1}^{r}\left(\sum_{k=0}^{q-1}(m\alpha)^k\right)^2 2^{r-q}\right)\exp\left(\lambda c\sum_{q=0}^{r}(m\alpha)^q\right),
\end{aligned}
$$

where the last inequality was obtained using hypothesis **(H1)**. In the sequel, c_0, c_1 and c_2 will denote some positive constants which depend on α, m, and c. They may differ from one line to another. For $m\alpha \neq 1$ and $m\alpha \neq \sqrt{2}$, we deduce from the foregoing and from (4.4) that

$$
\begin{aligned}
\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f)>\delta\right) &\le \exp(-\lambda\delta t_r)\exp\left(\frac{\lambda c(1-(m\alpha)^{r+1})}{1-m\alpha}\right)\\
&\quad\times\exp\left(\frac{c^2(2+m\alpha)^2\lambda^2}{2(1-m\alpha)^2}\left((2^r-1)-\frac{2m\alpha(2^r-(m\alpha)^r)}{2-m\alpha}\right.\right.\\
&\qquad\qquad\qquad\qquad\left.\left.+\frac{(m\alpha)^2(2^r-(m^2\alpha^2)^r)}{2-(m\alpha)^2}\right)\right)\\
&\le \exp\left(-\lambda\delta t_r+\frac{c^2(2+m\alpha)^2\lambda^2}{2(m\alpha-1)^2}\left((2^r-1)+\frac{(m\alpha)^2(2^r-(m^2\alpha^2)^r)}{2-(m\alpha)^2}\right)\right)\\
&\qquad\qquad\times\exp\left(\frac{\lambda c(1-(m\alpha)^{r+1})}{1-m\alpha}\right).
\end{aligned}
$$

Taking $\lambda = \dfrac{\delta t_r (m\alpha - 1)^2}{c^2(2+m\alpha)^2\left((2^r-1)+\frac{(m\alpha)^2(2^r-(m^2\alpha^2)^r)}{2-(m\alpha)^2}\right)}$, we are led to

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \leq \exp\left(-\frac{\delta^2(1-m\alpha)^2 t_r^2}{2c^2(2+m\alpha)^2\left(2^r-1+\frac{(m\alpha)^2(2^r-(m^2\alpha^2)^r)}{2-(m\alpha)^2}\right)}\right)$$
$$\times \exp\left(\frac{\delta(1-m\alpha)^2 t_r}{c(2+m\alpha)^2\left(2^r-1+\frac{(m\alpha)^2(2^r-(m^2\alpha^2)^r)}{2-(m\alpha)^2}\right)} \times \frac{1-(m\alpha)^{r+1}}{1-m\alpha}\right).$$

Now, the rest of the proof divides into five cases.

• If $m\alpha < 1$, then, for all $r \in \mathbb{N}$, $(m\alpha)^{r+1} - 1 \leq m\alpha - 1$ and $2^r - (m\alpha)^{2r} < 2^r$. We then deduce that

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \leq \exp(c_2\delta)\exp\left(-c_2\delta^2(m^2/2)^{r+1}\right).$$

• If $1 < m\alpha < \sqrt{2}$, then we have

$$\begin{aligned}\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) &\leq \exp\left(-c_1\delta^2(m^2/2)^{r+1}\right)\exp\left(c_2\delta\frac{(m\alpha)^{r+1}-1}{m\alpha-1}\right)\\ &\leq \exp\left(-\delta c_2(m^2/2)^{r+1}(\delta - c_0\alpha^{r+1})\right).\end{aligned}$$

Now, for all $r \in \mathbb{N}$ such that $r+1 > \log(\delta/2c_0)/\log(\alpha)$, we have $\delta - c_0\alpha^{r+1} > \delta/2$, in such a way that

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \leq \exp\left(\delta^2 c_2(m^2/2)^{r+1}\right).$$

• If $m\alpha > \sqrt{2}$, then for all $r \in \mathbb{N}$, $(m^2\alpha^2)^r > 2^r$. We then have

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \leq \exp\left(-c_2\delta\alpha^{-2r}(\delta - c_0\alpha^{r+1})\right).$$

Now for all $r \in \mathbb{N}$ such that $r+1 > \log(\delta/c_0)/\log(\alpha)$, we have

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \leq \exp\left(-\frac{c_2\delta^2}{\alpha^{2r}}\right).$$

• If $m\alpha = 1$, then

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \leq \exp\left(-\lambda\delta t_r + c_1 2^r\lambda^2\right)\exp\left(\lambda c(r+1)\right)$$

Taking $\lambda = \delta t_r / c_1 2^{r+1}$, we are led to

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \leq \exp\left(c_1\delta\frac{(r+1)t_r}{2^{r+1}}\right)\exp\left(-c_2\delta^2(m^2/2)^{r+1}\right).$$

• If $m\alpha = \sqrt{2}$, then

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \leq \exp\left(-\lambda\delta t_r + \lambda^2 c_1(r+1)2^r\right)\exp\left(\lambda c_1(\sqrt{2})^{r+1}\right).$$

Taking $\lambda = \delta t_r/(2c_1(r+1)2^r)$, we are led to

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \leq \exp\left(-\frac{c_2\delta}{r+1}\left(\frac{m^2}{2}\right)^{r+1}\left(\delta - c_0\left(\frac{\sqrt{2}}{m}\right)^{r+1}\right)\right).$$

Now, for all $r \in \mathbb{N}$ such that $r + 1 > \log(\delta/c_0)/\log(\sqrt{2}/m)$, we get

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \leq \exp\left(-\frac{c_2\delta^2}{r+1}\left(\frac{m^2}{2}\right)^{r+1}\right).$$

This ends the proof for $\mathbb{H}_r = \mathbb{T}_r$.

4.2. **Proof of Theorem 2.3.** Let $f \in \mathcal{B}_b(S)$ such that $\langle\mu, f\rangle \neq 0$. Once again, we are going to study successively $\widetilde{M}_{\mathbb{G}_r^*}(f)$ and $\widetilde{M}_{\mathbb{T}_r^*}(f)$.

Step 1. Let us first deal with $\widetilde{M}_{\mathbb{G}_r^*}(f)$. Set $g = f - \langle\mu, f\rangle$. Then, $\langle\mu, g\rangle = 0$ and

$$\widetilde{M}_{\mathbb{G}_r^*}(f) = \widetilde{M}_{\mathbb{G}_r^*}(g) + (|\mathbb{G}_r^*|/m^r)\langle\mu, f\rangle.$$

We have

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) - \langle\mu, f\rangle W > \delta\right) \leq \mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(g) > \delta/2\right) + \mathbb{P}\left(\left|\frac{|\mathbb{G}_r^*|}{m^r} - W\right| > \frac{\delta}{2|\langle\mu, f\rangle|}\right). \tag{4.6}$$

As $\langle\mu, g\rangle = 0$, the previous computations (proof of Theorem 2.2) give us some bound for the first term of right hand of (4.6), similar to those obtain in Theorem 2.2. Now, under hypothesis **(H3)**, we deduce, from [2] Theorem 5, that

$$\mathbb{P}\left(\left|\frac{|\mathbb{G}_r^*|}{m^r} - W\right| > \frac{\delta}{2|\langle\mu, f\rangle|}\right) \leq c_2 \exp\left(-c_3\delta^{2/3}m^{r/3}\right),$$

and this ends the proof of Theorem 2.3 when $\mathbb{H}_r = \mathbb{G}_r$.

Step 2. Let us look at $\widetilde{M}_{\mathbb{T}_r^*}(f)$. For $f \in \mathcal{B}_b(S)$, set $g = f - \langle\mu, f\rangle$. Then, $\langle\mu, g\rangle = 0$ and

$$\widetilde{M}_{\mathbb{T}_r^*}(f) = \widetilde{M}_{\mathbb{T}_r^*}(g) + (|\mathbb{T}_r^*|/t_r)\langle\mu, f\rangle.$$

We have

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) - \langle\mu, f\rangle W > \delta\right) \leq \mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(g) > \delta/2\right) + \mathbb{P}\left(\left|\frac{|\mathbb{T}_r^*|}{t_r} - W\right| > \frac{\delta}{2|\langle\mu, f\rangle|}\right). \tag{4.7}$$

Since $\langle\mu, g\rangle = 0$, the first term of the right hand of (4.7) can be bounded as in the previous computations (proof of Theorem 2.2). Under additional hypothesis **(H3)**, we have, from [2] Theorem 5,

$$\begin{aligned}\mathbb{P}\left(\left|\frac{|\mathbb{T}_r^*|}{t_r} - W\right| > \frac{\delta}{2|\langle\mu, f\rangle|}\right) &\leq \sum_{q=0}^{r}\mathbb{P}\left(\frac{m^q}{t_r}\left|\frac{|\mathbb{G}_q^*|}{m^q} - W\right| > \frac{\delta}{2(r+1)|\langle\mu, f\rangle|}\right)\\ &= \sum_{q=0}^{r}\mathbb{P}\left(\left|\frac{|\mathbb{G}_q^*|}{m^q} - W\right| > \frac{\delta t_r}{2(r+1)|\langle\mu, f\rangle|m^q}\right)\end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{q=0}^{r} c_2 \exp\left(-c_3 \delta^{2/3} \left(\frac{t_r^2}{(r+1)m^q}\right)^{1/3}\right) \\
&\leq c_2 \exp\left(-c_3 \delta^{2/3} \left(\frac{t_r}{(r+1)^2}\right)^{1/3}\right) \Big(1+o(1)\Big),
\end{aligned}
$$

and this ends the proof of Theorem 2.3 when $\mathbb{H}_r = \mathbb{T}_r$.

4.3. **Proof of Theorem 2.5.** Let $f \in \mathcal{B}_b(S)$. Without loss of generality, we assume that $\langle \mu, f\rangle = 0$. Otherwise, we take $f - \langle \mu, f\rangle$. For all $\delta > 0$, for all $a > 0$ and for all $b > 0$ such that $b < a/(\delta+1)$, we have

$$
\begin{aligned}
\mathbb{P}\left(\overline{M}_{\mathbb{H}_r^*}(f) > \delta | W \geq a\right) &= \mathbb{P}\left(\overline{M}_{\mathbb{H}_r^*}(f) > \delta, \frac{|\mathbb{H}_r^*|}{h_r} > b | W \geq a\right) \\
&\quad + \mathbb{P}\left(\overline{M}_{\mathbb{H}_r^*}(f) > \delta, \frac{|\mathbb{H}_r^*|}{h_r} \leq b | W \geq a\right) \\
&= \frac{1}{\mathbb{P}(W \geq a)} \Bigg(\mathbb{P}\left(\overline{M}_{\mathbb{H}_r^*}(f) > \delta, \frac{|\mathbb{H}_r^*|}{h_r} > b, W \geq a\right) \\
&\quad + \mathbb{P}\left(\overline{M}_{\mathbb{H}_r^*}(f) > \delta, \frac{|\mathbb{H}_r^*|}{h_r} \leq b, W \geq a\right)\Bigg) \\
&\leq p_a \mathbb{P}\left(\widetilde{M}_{\mathbb{H}_r^*}(f) > \delta b\right) + p_a \mathbb{P}\left(\left|\frac{|\mathbb{H}_r^*|}{h_r} - W\right| > W - b, W \geq a\right) \\
&\leq p_a \mathbb{P}\left(\widetilde{M}_{\mathbb{H}_r^*}(f) > \delta b\right) + p_a \mathbb{P}\left(\left|\frac{|\mathbb{H}_r^*|}{h_r} - W\right| > \delta b\right),
\end{aligned}
$$

where $p_a = \mathbb{P}(W \geq a)^{-1}$. Now, the first term of the last inequality can be bounded as in Theorem 2.2, and the second term is bounded as in the **step 1** and and **step 2** of the proof of Theorem 2.3. This ends the proof.

4.4. **Proof of Theorem 2.6.** Let $f \in \mathcal{B}_b(S^3)$.

Step 1. Let us first deal with $\widetilde{M}_{\mathbb{G}_r^*}(f)$. Assume that $\langle \mu, P^* f\rangle = 0$. By Chernoff inequality, we have for all $\delta > 0$ and for all $\lambda > 0$,

$$
\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) > \delta\right) \leq \exp(-\lambda \delta m^r)\, \mathbb{E}\left[\exp\left(\lambda \sum_{i \in \mathbb{G}_r^*} f(\Delta_i)\right)\right].
$$

Conditioning by $\mathcal{F}_r$, and using, conditional independence of triplets $\{\Delta_i, i \in \mathbb{G}_r\}$ with respect to $\mathcal{F}_r$, Azuma-Bennet-Hoeffding inequality and **(H2)**, we obtain

$$
\begin{aligned}
&\mathbb{E}\left[\exp\left(\lambda \sum_{i \in \mathbb{G}_r^*} f(\Delta_i)\right)\right] \\
&\quad = \mathbb{E}\left[\exp\left(\lambda \sum_{i \mathbb{G}_r^*} P^* f(X_i)\right) \mathbb{E}\left[\exp\left(\lambda \sum_{i \in \mathbb{G}_r^*} (f(\Delta_i) - P^* f(X_i))\right) \Big| \mathcal{F}_r\right]\right]
\end{aligned}
$$

$$= \mathbb{E}\left[\exp\left(\lambda \sum_{i\mathbb{G}_r^*} P^* f(X_i)\right) \prod_{i\in\mathbb{G}_r^*} \mathbb{E}\left[\exp\left(\lambda(f(\Delta_i) - P^* f(X_i))\right) \Big| \mathcal{F}_r\right]\right]$$
$$\leq \exp\left(2\lambda^2 \|f\|_\infty c_1 m^r\right) \mathbb{E}\left[\exp\left(\lambda \sum_{i\in\mathbb{G}_r^*} P^* f(X_i)\right)\right].$$

We control the last expectation as in the **Step 1** of the proof of Theorem 2.2, apply to P^*f. Next, we get the result discussing as in the proof of Theorem 2.2.
If $\langle \mu, P^* f\rangle \neq 0$, we set $g = f - \langle \mu, P^* f\rangle$. Then, we have

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) - \langle \mu, P^* f\rangle W > \delta\right) \leq \mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(g) > \delta/2\right) + \mathbb{P}\left(\left|\frac{\mathbb{G}_r^*}{m^r} - W\right| > \delta/2 |\langle \mu, P^* f\rangle|\right). \tag{4.8}$$

The first term of the right hand of (4.8) can be bounded as previously since $\langle \mu, P^* g\rangle = 0$. The second term can be bounded as in **Step 1** of the proof of Theorem 2.2. This ends the proof for $\widetilde{M}_{\mathbb{G}_r^*}(f)$.

Step 2. Let us now treat $\widetilde{M}_{\mathbb{T}_r^*}(f)$. First, we assume that $\langle \mu, P^* f\rangle = 0$. For all $\delta > 0$, we have

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \leq \mathbb{P}\left(\frac{1}{t_r}\sum_{i\in\mathbb{T}_r^*}(f(\Delta_i) - P^* f(X_i)) > \delta/2\right) + \mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(P^* f) > \delta/2\right).$$

By chernoff inequality, we have for all $\lambda > 0$,

$$\mathbb{P}\left(\frac{1}{t_r}\sum_{i\in\mathbb{T}_r^*}(f(\Delta_i) - P^* f(X_i)) > \delta/2\right) \leq \exp\left(-\frac{\lambda\delta t_r}{2}\right) \times \mathbb{E}\left[\exp\left(\lambda \sum_{i\in\mathbb{T}_r^*}(f(\Delta_i) - P^* f(X_i))\right)\right]$$

Conditioning successively with respect to $(\mathcal{F}_q)_{0\leq q\leq r}$, using conditional independence of triplets $\{\Delta_i, i\in\mathbb{G}_q\}$ with respect to $\mathcal{F}_q$ and applying successively Azuma-Bennet-Hoeffding inequality and the fact that $|\mathbb{G}_q^*| \leq |\mathbb{G}_q|$ for all $q\in\{0,\cdots,r\}$, we get

$$\mathbb{E}\left[\exp\left(\lambda \sum_{i\in\mathbb{T}_r^*}(f(\Delta_i) - P^* f(X_i))\right)\right]$$
$$= \mathbb{E}\left[\exp\left(\lambda \sum_{i\in\mathbb{T}_{r-1}^*}(f(\Delta_i) - P^* f(X_i))\right) \mathbb{E}\left[\exp\left(\lambda \sum_{i\in\mathbb{G}_r^*}(f(\Delta_i) - P^* f(X_i))\right) \Big| \mathcal{F}_r\right]\right]$$
$$= \mathbb{E}\left[\exp\left(\lambda \sum_{i\in\mathbb{T}_{r-1}^*}(f(\Delta_i) - P^* f(X_i))\right) \prod_{i\in\mathbb{G}_r^*} \mathbb{E}\left[\exp\left(\lambda(f(\Delta_i) - P^* f(X_i))\right) \Big| \mathcal{F}_r\right]\right]$$

$$\leq \exp\left(2\lambda^2\|f\|_\infty^2|\mathbb{G}_r|\right)\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}^*_{r-1}}(f(\Delta_i)-P^*f(X_i))\right)\right]$$
$$\vdots$$
$$\leq \exp\left(2\lambda^2\|f\|_\infty^2|\mathbb{T}_r|\right).$$

Next, optimizing on λ, we obtain

$$\mathbb{P}\left(\frac{1}{t_r}\sum_{i\in\mathbb{T}^*_r}(f(\Delta_i)-P^*f(X_i))>\delta/2\right)\leq\exp\left(-c_1\delta^2\left(\frac{m^2}{2}\right)^{r+1}\right),$$

for some positive constant c_1. The term $\mathbb{P}\left(\widetilde{M}_{\mathbb{T}^*_r}(P^*f)>\delta/2\right)$ can be bounded as in the proof of Theorem 2.2, and this ends the proof when $\langle\mu,P^*f\rangle=0$. On the other hand, if $\langle\mu,P^*f\rangle\neq 0$, we have

$$\widetilde{M}_{\mathbb{T}^*_r}(f)-\langle\mu,P^*f\rangle W=\widetilde{M}_{\mathbb{T}^*_r}(g)+\left(\frac{|\mathbb{T}^*_r|}{t_r}-W\right)\langle\mu,P^*f\rangle.$$

We then proceed as for (4.8), and this ends the proof for $\widetilde{M}_{\mathbb{T}^*_r}(f)$.

Step 3. Eventually, we bound $\mathbb{P}\left(\overline{M}_{\mathbb{H}^*_r}(f)>\delta-\langle\mu,P^*f\rangle>\delta\right)$, using **Step 1** and **Step 2**, as in the proof of Theorem 2.5.

4.5. **Proof of Proposition 3.1.** We are going to treat $\widehat{\alpha}_0^n-\alpha_0$. Deviation inequalities for $\widehat{\alpha}_1^n-\alpha_1$, $\widehat{\beta}_\eta^n-\beta_\eta$, $\widehat{\alpha}_\eta^{\prime n}-\alpha_\eta^{\prime}$, $\widehat{\beta}_\eta^{\prime n}-\beta_\eta^{\prime}$, $\eta\in\{0,1\}$, can be treated in the same way. Recalling that the state space of the process $(X_i, i\in\mathbb{T}^*)$, denoted by S, is assumed to be a compact subset of $\mathbb{R}$.

Let g_1, g_2, h_1 and h_2 the functions defined on S^3 respectively by

$$\begin{aligned}
g_1(x,y,z)&=(xy-x(\alpha_0x+\beta_0))\,\mathbf{1}_{S^3}(x,y,z),\\
g_2(x,y,z)&=(y-\alpha_0x-\beta_0))\,\mathbf{1}_{S^3}(x,y,z),\\
h_1(x,y,z)&=x\mathbf{1}_{S^3}(x,y,z),\\
h_2(x,y,z)&=x^2\mathbf{1}_{S^3}(x,y,z).
\end{aligned}$$

It is easy to see that $P^*g_1(x)=0$, $P^*g_2(x)=0$, $P^*h_1(x)=p_{1,0}x$ and $P^*h_2(x)=p_{1,0}x^2$ where P^* denote the transition kernel associated to the BAR(1) process with missing data. With these notations, we can rewrite $\widehat{\alpha}_0^n-\alpha_0$ as

$$\widehat{\alpha}_0^n-\alpha_0=\frac{|\mathbb{T}^*_n|^{-1}|\mathbb{T}^{1,0}_n|\left(|\mathbb{T}^*_n|^{-1}\sum_{i\in\mathbb{T}^*_n}g_1(\Delta_i)\right)}{B_n}
-\frac{\left(|\mathbb{T}^*_n|^{-1}\sum_{i\in\mathbb{T}^*_n}h_1(\Delta_i)\right)\left(|\mathbb{T}^*_n|^{-1}\sum_{i\in\mathbb{T}^*_n}g_2(\Delta_i)\right)}{B_n},$$

where $B_n=|\mathbb{T}^*_n|^{-1}|\mathbb{T}^{1,0}_n|\left(|\mathbb{T}^*_n|^{-1}\sum_{i\in\mathbb{T}^*_n}h_2(\Delta_i)\right)-\left(|\mathbb{T}^*_n|^{-1}\sum_{i\in\mathbb{T}^*_n}h_1(\Delta_i)\right)^2.$

Recalling (1.8), we then have for all $\delta > 0$ and $a > 0$

$$\mathbb{P}\left(|\widehat{\alpha}_0^n - \alpha_0| > \delta | W \geq a\right) \leq \mathbb{P}\left(\frac{|\mathbb{T}_n^*|^{-1}|\mathbb{T}_n^{1,0}||\overline{M}_{\mathbb{T}_n^*}(g_1)|}{|B_n|} > \frac{\delta}{2}\Big| W \geq a\right) \qquad (4.9)$$
$$+\mathbb{P}\left(\frac{|\overline{M}_{\mathbb{T}_n^*}(h_1)||\overline{M}_{\mathbb{T}_n^*}(g_2)|}{|B_n|} > \frac{\delta}{2}\Big| W \geq a\right).$$

For the first term of the right hand of (4.9), since $|\mathbb{T}_n^*|^{-1}|\mathbb{T}_n^{1,0}| \leq 1$, we have for all $\gamma > 0$

$$\mathbb{P}\left(\frac{|\mathbb{T}_n^*|^{-1}|\mathbb{T}_n^{1,0}||\overline{M}_{\mathbb{T}_n^*}(g_1)|}{|B_n|} > \frac{\delta}{2}\Big| W \geq a\right) \leq \mathbb{P}\left(|B_n| < \gamma | W \geq a\right)$$
$$+\mathbb{P}\left(|\overline{M}_{\mathbb{T}_n^*}(g_1)| > \frac{\delta\gamma}{2}\Big| W \geq a\right).$$

Notice that

$$B_n - (p_{1,0}^2\mu_2 - p_{1,0}^2\mu_1^2) = p_{1,0}\mu_2\left(\frac{|\mathbb{T}_n^{1,0}|}{|\mathbb{T}_n^*|} - p_{1,0}\right) + \frac{|\mathbb{T}_n^{1,0}|}{|\mathbb{T}_n^*|}\overline{M}_{\mathbb{T}_n^*}(h_2 - p_{1,0}\mu_2)$$
$$- \left(\overline{M}_{\mathbb{T}_n^*}(h_1 - p_{1,0}\mu_1)\right)^2 - 2p_{1,0}\mu_1\overline{M}_{\mathbb{T}_n^*}(h_1 - p_{1,0}\mu_1)$$

and

$$\{|B_n| < \gamma\} \subset \left\{|B_n - (p_{1,0}^2\mu_2 - p_{1,0}^2\mu_1^2)| > |p_{1,0}^2\mu_2 - p_{1,0}^2\mu_1^2| - \gamma\right\}.$$

We then have for all $0 < \gamma < \dfrac{2|p_{1,0}^2\mu_2 - p_{1,0}^2\mu_1^2|}{2+\delta}$,

$$\mathbb{P}\left(\frac{|\mathbb{T}_n^*|^{-1}|\mathbb{T}_n^{1,0}||\overline{M}_{\mathbb{T}_n^*}(g_1)|}{|B_n|} > \frac{\delta}{2}\Big| W \geq a\right)$$
$$\leq \mathbb{P}\left(|B_n| < \gamma | W \geq a\right) + \mathbb{P}\left(|\overline{M}_{\mathbb{T}_n^*}(g_1)| > \frac{\delta\gamma}{2}\Big| W \geq a\right)$$
$$\leq \mathbb{P}\left(|B_n - (p_{1,0}^2\mu_2 - p_{1,0}^2\mu_1^2)| > \frac{\gamma\delta}{2}\Big| W \geq a\right) + \mathbb{P}\left(|\overline{M}_{\mathbb{T}_n^*}(g_1)| > \frac{\delta\gamma}{2}\Big| W \geq a\right)$$
$$\leq \mathbb{P}\left(|\overline{M}_{\mathbb{T}_n^*}(g_1)| > \frac{\delta\gamma}{2}\Big| W \geq a\right) + \mathbb{P}\left(|\overline{M}_{\mathbb{T}_n^*}(h_2 - p_{1,0}\mu_2)| > \frac{\gamma\delta}{8}\Big| W \geq a\right)$$
$$+ \mathbb{P}\left(|\overline{M}_{\mathbb{T}_n^*}(h_1 - p_{1,0}\mu_1)| > \frac{\sqrt{\gamma\delta}}{2\sqrt{2}}\Big| W \geq a\right) + \mathbb{P}\left(\left|\frac{|\mathbb{T}_n^{1,0}|}{|\mathbb{T}_n^*|} - p_{1,0}\right| > \frac{\delta\gamma}{8p_{1,0}\mu_2}\Big| W \geq a\right)$$
$$+ \mathbb{P}\left(|\overline{M}_{\mathbb{T}_n^*}(h_1 - p_{1,0}\mu_1)| > \frac{\gamma\delta}{16p_{1,0}\mu_1}\Big| W \geq a\right).$$

From [13], Section 5, we have

$$\mathbb{P}\left(\left|\frac{|\mathbb{T}_n^{1,0}|}{|\mathbb{T}_n^*|} - p_{1,0}\right| > \frac{\delta\gamma}{8p_{1,0}\mu_2}\Big| W \geq a\right) = \mathbb{P}\left(\left|\frac{1}{|\mathbb{T}_n^*|}\sum_{j=1}^{|\mathbb{T}_n^*|}(T_j - p_{1,0})\right| > \frac{\delta\gamma}{8p_{1,0}\mu_2}\Big| W \geq a\right), \qquad (4.10)$$

where $(T_j)_{j\geq 1}$ is a sequence of i.i.d. Bernoulli random variables such that

$$p_{1,0} = \mathbb{P}(T_j = 1) = 1 - \mathbb{P}(T_j = 0).$$

To majorize the right hand side of (4.10), we use exactly the same ideas that for the proof of Theorem 2.5 and Step 2 of the proof of Theorem 2.2.

For the second term of the right hand of (4.9), we have

$$\mathbb{P}\left(\frac{|\overline{M}_{\mathbb{T}_n^*}(h_1)||\overline{M}_{\mathbb{T}_n^*}(g_1)|}{|B_n|} > \frac{\delta}{2}\Big|W \geq a\right) \leq \mathbb{P}\left(\frac{|\overline{M}_{\mathbb{T}_n^*}(g_2)|}{|B_n|} > \frac{\delta}{4p_{1,0}\mu_1}\Big|W \geq a\right)$$
$$+ \mathbb{P}\left(\frac{|\overline{M}_{\mathbb{T}_n^*}(g_2)|}{|B_n|} > \frac{\sqrt{\delta}}{2}\Big|W \geq a\right) + \mathbb{P}\left(|\overline{M}_{\mathbb{T}_n^*}(h_1 - p_{1,0}\mu_1)| > \frac{\sqrt{\delta}}{2}\Big|W \geq a\right)$$

Now, the first and the second term of the right hand of the last inequality can be treated as the first term of the right hand of (4.9).

Finally, to get the result, just apply Theorem 2.6 to functions g_1, g_2, h_1 and h_2.

Acknowledgments. The author thanks its advisor Pr. Arnaud GUILLIN for all his advices and suggestions during the preparation of this work.

References

[1] Adamczak, R., and Miłoś, P. CLT for ornstein-uhlenbeck branching particle system. *arXiv:1111.4559* (2011).
[2] Athreya, K. Large deviations for branching processes-i. single type case. *Ann. Appl. Probab. 5, No. 3* (1994), 779–790.
[3] Athreya, K., and Ney, P. *Branching Process*. Springer, Berlin, 1972.
[4] Azuma, K. Weighted sums of certain dependent random variables. *Tôhoku Math. J 19, No. 3* (1967), 357–367.
[5] Bennett, G. Probability inequalities for sum of independant random variables. *Journal of the American Statistical Association 57, No. 297* (1962), 33–45.
[6] Bercu, B., De Saporta, B., and Gégout-Petit, A. Asymptotic analysis for bifurcating autoregressive processes via a martingale approach. *Electronic. J. Probab. 14* (2009), 2492–2526.
[7] Bitseki Penda, S. V., and Djellout, H. Deviation inequalities and moderate deviations for estimators of parameters in bifurcating autoregressive models. *arXiv:1204.2355v1. Soumis* (2012).
[8] Bitseki Penda, S. V., Djellout, H., and Guillin, A. Deviation inequalities, moderate deviations and some limit theorems for bifurcating Markov chains with application. *Preprint* (2011).
[9] Cowan, R., and Staudte, R. G. The bifurcating autoregressive model in cell lineage studies. *Biometrics 42* (1986), 769–783.
[10] De Saporta, B., Gégout-Petit, A., and Marsalle, L. Asymmetry tests for bifurcating autoregressive processes with missing data. *preprint*, 2011.
[11] De Saporta, B., Gégout-Petit, A., and Marsalle, L. Parameters estimation for asymmetric bifurcating autoregressive processes with missing data. *Electronic Journal of Statistics 5* (2011), 1313–1353.
[12] Delmas, J. F., and Marsalle, L. Detection of cellular aging in Galton-Watson process. *Stochastic Processes and their Applications 120* (2010), 2495–2519.
[13] Dion, J., and Yanev, N. M. Limit theorems and estimation theory for branching processes with increasing random number of ancestors. *J. Appl. Probab. 34* (1997), 309–327.
[14] Guyon, J. Limit theorems for bifurcating Markov chains. application to the detection of cellular aging. *Ann. Appl. Probab. Vol. 17, No. 5-6* (2007), 1538–1569.

[15] Guyon, J., Bize, A., Paul, G., Stewart, E., Delmas, J., and Taddéi, F. Statistical study of cellular aging. *CEMRACS 2004 Proceedings, ESAIM Proceedings 14* (2005), 100–114.
[16] Hoeffding, W. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association 58, No. 301* (1963), 13–30.
[17] Stewart, E. J., Madden, R., Paul, G., and Taddéi, F. Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biol 3(2)* (2005), e45.

Valère BITSEKI PENDA, Laboratoire de Mathématiques, CNRS UMR 6620, Université Blaise Pascal, avenue des Landais 63177 Aubière.

E-mail address: Valere.Bitsekipenda@math.univ-bpclermont.fr